

## TYPO3 Core - Bug #95007

### Checking the file checksum on export or import may fail due to sanitizers

2021-08-26 13:35 - Alexander Nitsche

<b>Status:</b>	New	<b>Start date:</b>	2021-08-26
<b>Priority:</b>	Should have	<b>Due date:</b>	
<b>Assignee:</b>		<b>% Done:</b>	0%
<b>Category:</b>	Import/Export (T3D)	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	next-patchlevel	<b>Complexity:</b>	
<b>TYPO3 Version:</b>	11	<b>Is Regression:</b>	
<b>PHP Version:</b>		<b>Sprint Focus:</b>	
<b>Tags:</b>			

#### Description

The calculated file checksums of the exported sys\_file records in a TYPO3 export dump - made via EXT:impexp - may not match the actual checksum of the exported file, since the checksum is calculated before the file is copied to the export folder, and the checksum may change during this copying process. The copy process is done by GeneralUtility::upload\_copy\_move, which applies arbitrary sanitizers (e.g. SVG sanitizer) under the hood. One solution would be to recalculate the checksum *after* the file is copied.

The same applies for the import: During import the sanitizers may change the file content and make the final checksum test fail. It has to be checked, if this test should rather raise a notice than an error on failure, or if the test can be modified to still ensure validity.

#### Related issues:

Related to TYPO3 Core - Story #92526: Concept for improved EXT:impexp

Accepted

2020-10-09

#### History

##### #1 - 2021-08-27 17:49 - Oliver Hader

Taken from a Slack conversation:

Alexander Nitsche

so, the import + \*sanitizer resolves currently to:

- (1) copy file-to-be-imported to temporary folder by GeneralUtility::writeFile (if file is in the data.xml) or GeneralUtility::copyDirectory (if file is in extra folder data.xml.files) => this probably does not change the file and hash
- (2) check if there is already a file on the system with the same identifier and hash as the file-to-be-imported => this check fails now unintentionally if the file on system has not been sanitized yet and the file-to-be-imported has - or vice-versa
- (3) copy file to system by ResourceStorage::addFile() which probably sanitizes the file
- (4) check if the final file hash matches the related sys\_file record hash and throw an error if not: this happens currently if an unsanitized file got sanitized during storage->addFile()

suggestion: we could relax the final check (4) and move it up to after (1) and check if the file hash in the temporary folder matches the related sys\_file record hash and if so, this is fine too.

missing: what to do about check (2): how can we determine reliably, that an existing file on the system matches the file-to-be-imported if one could be sanitized and the other not?

Oliver Hader

The file hash given in XML export files is used to ensure integrity - it is not a security feature.

Thus, checking integrity after step (1) seems to be fine - since that's the low-level processing of files.

Step (2) cannot be solved in a generic way, without adding special behavior for e.g. SVG files. Besides that, the import process is based on the assumption, that existing files should be "reused" (referenced).

If that principle would be applied consequently in TYPO3, it would not be possible in filelist module to copy a file from path A to path B - thus, there are scenarios where duplicates are expected - from my POV it's not the job of the import process to implicitly resolve that.

Actually a dedicated task/job/migration-wizard should be used to identify similar files (based on their content-hash) and adjust corresponding references, reducing redundant information.

##### #2 - 2021-09-17 13:43 - Alexander Nitsche

- Subject changed from *Recalculate file checksum after copy to export folder* to *Checking the file checksum on export or import may fail due to sanitizers*

- Description updated

**#3 - 2021-09-17 13:43 - Alexander Nitsche**

- Related to Story #92526: Concept for improved EXT:impexp added